SURVEY: MACHINE LEARNING AND CHALLENGES

ANSAR SHEIKH[a1] AND S.K. PATHAK[b]

[ab] Department of Computer Science & Engineering, Dr. K.N. Modi University, Newai, Rajasthan, India

ABSTRACT

Big amount of data is available everywhere. Some of them are useful. It is important to analyze this data in term of usefulness which help to develop using training data and some data are used to test on developed ML model. Machine learning is the integral part of artificial intelligent which is used to design algorithm based on training data. Machine learning has many applications also have many challenges in some distinct area such as discriminative learning, ML bottlenecks and so on. This paper presents the machine learning with some crucial challenges in few different areas.

KEYWORDS: Machine Learning, Classification, Linear Regression, Clustering and Association

Machine learning is concept to learn the computer from the day today's example and past experiences and does not required to program machine. It provides ability of learning to computer which predicts the things as human being determined. ML is modeled with large data i.e. training data, to predict approximately accurate output is to be called test data. ML is fun. ML products trend in market. ML is difficult to model with lack of knowledge of artificial intelligent. Because of this reason, ML has many challenges.

ML is divided into 3 types as supervised learning, unsupervised learning, and reinforcement learning.

Supervised learning: The dataset acts as teacher incorporate with the training data and output with machine learning rule. Model continue the training until get the accurate prediction. Machine learning and Challenges are discussed in Section II,and III.

MACHINE LEARNING

ML is a class of algorithm which allows machine to learn from the experience and example without being explicitly programmed. So feed training data to generic algorithm which build a logistic model. In tradition system, we give input and program to a machine learning which generate the output. But in ML which needs to give training data and output as input to algorithm and produce prediction as output. New data given to machine leaning model which give the accurate prediction then the algorithm is deployed otherwise again train it. Algorithm is divided into training and testing. Training data and output are given to algorithm as input in training phase and testing with new data occurred in testing phase.

ML can be applicable in various areas such as Google map, social networking, taxi navigation apps, Facebook, Netflix and trading market and many more. Machine learner must require more study to apply before implementing machine learning model. ML has three categories as follows:

Supervised Machine Learning

Supervised Learning where you can consider learning is guided by the teacher who is dataset and apply rule to train the model. New input is given to trained model which predict accurately. For example, spammed and non-spammed emails are fed ML algorithm to indicate mail is spam or non-spam by generating accurate prediction rule[2]. SLM has four model i.e. Regression ,Decision tree , Random forest and Classification.

Regression

Regression is in which training data is in the form of values feed to algorithm and provide a prediction value. Linear and polynomial are the type of regression. Linear regression is used to for predicting unknown value of dependent variable from a known variables (independent variable).Taking two features drew a predicted line of regression in a graph by calculating approximate prediction by a model. Regression line is simply a line that best fits the data.

---

[1]**Corresponding author**
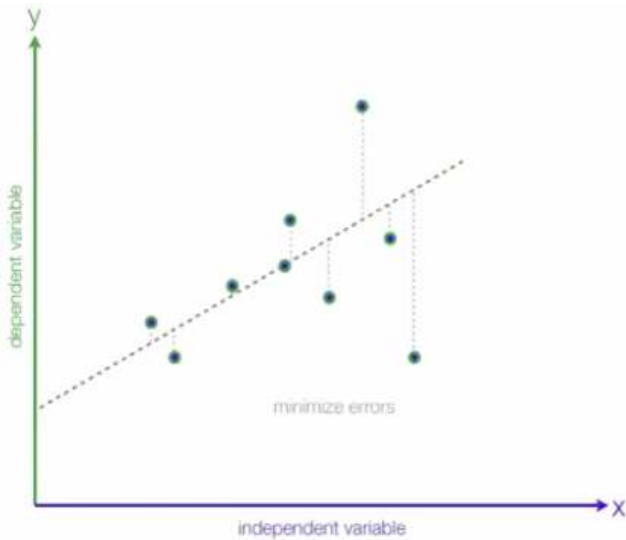
Model using large number of feature the will get more accuracy.

In polynomial, we can use correlated data for the prediction.

Decision Tree

It is tree like structure in which internal node represents test on a attribute. Each branch represents test outcome. Each leaf node is label. Path from root to leaf is called classification rule.

Random Forest

Random forest is ensemble classifier using many decision tree model. It has higher accuracy the decision tree

Classification

A classification problem is when the output variable is either "yes" or "no", "pass" and "fail". Classification model reaches to conclusion from observed values. Minimum one or more inputs a classification model will try to predict one label is yes or no. Classification problem are predicting person age from their handwriting, predicting monsoon will come next year [3].

UNSUPERVISED MACHINE LEARNING

Unsupervised learning is getting accurate prediction of trained data without having label in it Here the activity to make data in group which belongs to different kind of data group. The unsupervised learning is of following type.

Clustering

A clustering problem is grouping data inherently on the basis of the features such as customer group.

Association

An association rule learning problem is in which you find out that shows the part of date is belong to sum other category such as person is buying X, he will buy Y also as an assumption.

Reinforcement Learning

Reinforcement learning is worked in Machine Learning. It is about taking suitable action to maximize reward in a specific condition. It is used by various software and machines to find the best possible behavior or path it should take in a specific situation. Difference between supervised and reinforcement learning is that supervised learning haven't any answer, so model with corrected answer is trained. Whereas reinforcement agent has capability of acquiring experience by which determine the correct output when dataset is not available.

CHALLENGES IN ML

Reasoning Power

ML has not fully expert is in reasoning power, other than human trait. Algorithms available today are mainly oriented towards specific use-cases and are narrowed down when it comes to applicability. What is the reason behind to get outcome in this way For instance, if an image recognition algorithm identifies apples and oranges in a known situation, it is either apple or orange which is good or bad in state why is that fruit an apple or orange. Mathematically, all of this learning process can be explained by us, but from an algorithmic perspective, the innate property cannot be told by the algorithms or even us.

In other words, ML algorithm has lack of learning to identify the application requirement.

Contextual Limitation

If we consider the area of natural language processing (NLP), text and speech information are the means to understand languages by NLP algorithms. They may learn letters, words, sentences or even the syntax, but

where they fall back is the context of the language. Algorithms do not understand the context of the language used. A typical instance of Chinese room argument set by philosopher John Searle, which says that computer programs or algorithms grasp the idea merely by 'symbols' rather than the context given. (You are unable to get complete information by sitting in Chinese room).

So, ML does not have an overall idea of the situation. Mnemonic interpretations rather than thinking given a boundary to observe what is actually going on.

Scalability

On the important basis, we see ML implementations being deployed it is totally relied on data as well as its scalability. Data is growing at an enormous rate and has many forms which largely affect the scalability of an ML project. Algorithms will take care of updating new change in data. This is where ML looks the involvement of human for doing scalability and mostly solves the problem.

In addition, growing data has to be dealt the right way if shared on an ML platform which again needs examination through knowledge and intuition apparently lacked by current ML.

Regulatory Restriction for Data In ML

ML usually needs considerable amounts (in fact, massive) of data in stages such as training, cross-validation etc. Sometimes, data includes private as well as general information. This is place where it is difficult. Most tech companies have privatized data and these data are the ones which are actually useful for ML applications. But, there comes the risk of the wrong usage of data, especially in critical areas such as medical research, health insurance etc. Even though data are anonymised at times, it has the possibility of being vulnerable. Hence this is the reason regulatory rules are imposed heavily when it comes to using private data.

Internal Working of Deep Learning

This sub-field of ML is actually responsible for today's AI growth. Most powerful tool of ML is theory is useful. Deep Learning (DL) now powers applications such as voice recognition, image recognition and so on through artificial neural networks.

But, the DL's internal working is unknown and still in the process of solution. Advanced DL algorithms are complicated to understand the efficiently working for the many researchers. Millions of neurons that form the neural networks in DL increase abstraction at every level, which cannot be comprehended at all. Deep learning is called as black box since its internal agenda is unknown.

REFERENCES

https://www.geeksforgeeks.org/getting-started-machine-learning/

Sheena agra , Sachin ahuja "Machine Learning and its applications :a Review" 2017. International conference on big data analysis and computation intelligence", pp.57-60.

https://www.edureka.co/blog/what-is-machine-learning/#ML

Min, F,Hu.,& Zhu, W.(2014) "Feature selection from cost constraint. International Journal approximate reasoning 55(1),167-179.

Deep Residual Learning for Image Recognition, by He, K., Ren, S., Sun, J., & Zhang, X. (2016). CoRR, abs/1512.03385. (cited 1436 times, HIC: 137 , CV: 582).

A.L. Buczak and E. Guven, "A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection," IEEE Commun. Surv. Tutor., vol. 18, no. 2, pp. 1153–1176, 2016.